DOCUMENT RESUME

ED 458 294 TM 033 467

AUTHOR	Luecht, Richard M.
TITLE	Challenges of Web-Based Assessment.
PUB DATE	2001-04-00
NOTE	27p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Seattle, WA, April 11-13, 2001).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Computer Assisted Testing; Computer Interfaces; Educational Technology; Elementary Secondary Education; Measurement Techniques; *Test Construction; *Testing Problems; *World Wide Web
IDENTIFIERS	Test Security

ABSTRACT

This paper discusses six challenges relevant to Web-based testing. Some of these challenges are not specific to Web-based testing, but generalize to all computer-based testing. The challenges are: (1) security and using test centers for Web-based testing; (2) measuring complex skills and problem-solving tasks on the Web; (3) integrating modern item selection and test assembly algorithms; (4) storing and processing all relevant examinee response data; (5) the large-scale distribution of "high-bandwidth" tests (e.g., multimedia, high-density audio video, or images); and (6) optimal ergonomic design of Web-based testing interfaces. Considering each of these challenges raises questions about the future of Web-based testing and supports the need for better education across many sectors of the Web-based testing community about the technical aspects of psychometrics and high-stakes testing needs. In addition, there must be adherence to standards and principles of professional practice and science. (SLD)

Challenges of Web-Based Assessment[1]

Richard M. Luecht

University of North Carolina at Greensboro

April 2001

[1] Paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, WA, April 11-13, 2001. Invited syposium: *From the Printing Press to the World Wide Web: Implications for Knowledge Acquistion and Assessment.*

Introduction

The World Wide Web has enabled countless new technologies to emerge and will continue to do so. By the year 2003, the number of Internet users is predicted to climb to almost 200 million in the United States, and to 500 million worldwide. The 2000 *Campus Computing Survey* (Green, 2000) estimates that over three-fourths of higher education institutions now offer on-line services on their web sites ranging from e-mail services to web pages for courses. Everyone seems to be jumping on the *WWW* bandwagon. Testing is no exception.

For some applications, web-based testing and related online assessments offer the promise of rapid test authoring and deployment capabilities, 24 × 7 (twenty-four hours per day, seven days a week) access by examinees to testing, immediate feedback and scoring, prompt dissemination of paperless results and reports to the examinees, employers, teachers, or other users of scores, and a limited need for test administrators. The implication is that web-based testing is convenient, cost-effective and efficient. However, web-based testing (WBT) also raises concerns about unequal access of population subgroups to Internet technology, the security of systems, information protection and privacy, cheating and collaboration, and general fairness issues stemming from familiarity with WBT technologies. Beyond the commercial hype and promise of technological assessment capabilities, we need to critically evaluate the many purposes of testing on the web, what we are assessing, for whom, and under what conditions.

Wall (2000) nicely summarized four guidelines for using technology in testing and assessment: (1) conduct research to better understand the real advantages and disadvantages of technology use in specific contexts; (2) follow the assessment standards and policies of applicable professional associations and constituencies; (3) use "best practices" to ensure high quality assessment services; and (4) stay up-to-date on new research and topics related to assessment and technology. Keeping these guidelines in mind, this paper discusses six challenges relevant to web-based testing. Some of these challenges are not specific to WBT, per se, but generalize as well to computer-based testing (CBT).

Challenges of Web-Based Testing

This paper discusses six challenges for web-based testing: (1) security and using test centers for WBT; (2) measuring complex skills and problem-solving tasks on the web; (3) integrating modern item selection and test assembly algorithms; (4) storing and processing all relevant examinee response data, including "process information"; (5) large-scale distribution of "high-bandwidth" tests (e.g., multimedia, high-density audio-video, or images); and (6) optimal ergonomic design of web-based testing interfaces. There are obviously other topics that could be discussed. This list is not intended to be exhaustive.

Test Centers and the Security of On-Line Assessments

There may be a misconception by some that web-based testing—where the examinee conveniently sits at his or her home computer, orders up and takes a particular test, and gets the results immediately—will become the standard model for most computer-based tests in the future. Figure 1 presents a conceptual diagram of a client-server model for testing on the Internet. A simple schema can also verbally describe this model[2]. First, the Examinee (the client) connects to Web, receives authorization, and logs into the Examination Web Server. Second, the Examination Web Server authenticates the Examinee's workstation or PC and establishes a secure connection, usually using public key encryption. Third, software on the Examination Web Server selects question(s) and publishes the examination web page(s) and associated scripts; i.e., generates hypertext markup language (HTML) pages. [Note that actual examination units exchanged between the Examinee and the Server may consist of individual items, test sections or an entire test.] The Examination Web Server then *pushes* the web pages containing examination, sections, or questions to the Examinee, via the Internet connection. Fourth, the Examinee's web browser renders the examination (given its capabilities to display various test components and run scripts). The Examinee answers the questions using radio buttons, check boxes, text input boxes, etc. to record his or her responses. When the examination, test

[2] This schema and the conceptual diagram in Figure 1 are obviously over-simplified, but understanding the general flow between the Examinee and Server may prove useful.

section, or item is complete, the Examinee "submits" his or her responses to the Examination Server. Fifth, the Examination Web Server may score the information received up to that time and select addition items/sections to administer or stop, signifying the completion of the test. Sixth, upon receiving and building the complete data record of the examination, the Examination Web Server scores the examination, publishes a score report and pushes that report back to the Examinee, and possibly to others entitled to see the scores (e.g., an employer, teacher, school, or certifying agency). With only slight modification, this same model works for many web-based surveys.

This *examinee-as-client* WBT model demonstrates the convenience of scheduling and taking the test for the examinee, since (s)he merely had to sit in front of the computer, connect to the Internet, and complete the test. This model further suggests that the testing authority can globally deploy the examination anytime and anywhere an appropriate connection to the Internet can be established with examinees.

Unfortunately, this model is unlikely to work for most "high stakes" tests. There are many WBT applications for which this type of examinee-as-client model of testing is appropriate; high stakes testing is probably not one of them.

There are two broad classes of tests and assessments. One is amenable to web-based testing; the other is definitely "at risk" under standard applications of WBT. **Low-stakes** examinations are ideal for WBT. These tests include practice examinations, formative or diagnostic feedback examinations, self-assessments from training courses, course quizzes, and surveys. They tend to not require secure data exchange and involve measurement situations where decisions, interpretations, or other uses of the scores have few, if any, consequences for the examinee, the testing program or the testing authorities. In contrast, high stakes examinations include situations such as final course examinations, academic exit examinations, college or graduate school entrance tests, professional certification and licensure tests, job selection tests, and clinical psychological examinations.

It is important to realize that giving an examinee access to a high-stakes examination is identical to making it a "take-home" examination. In fact, with WBT, the examinee does not even have to take the exam home — the test is delivered directly to the client's personal computer. It seems improbable that employers seeking to screen potential examinees, colleges making admissions decisions, licensing or certification agencies entrusted with protecting the public, can naively invoke an "honor code" and trust examinees enough to self-administer high-stakes take-home examinations.

One of the major drawbacks of web-based testing, especially in unproctored environments, is security. That is, how can the testing authority (teacher, employment specialist, professional certifying body, licensing authority, testing organization, etc.) authenticate the identity of the examinee and that the performance submitted as the assessment is that examinee's performance? In proctored settings, examinees are often required to provide one or two forms of identification with a photo. How can the testing authority verify the integrity of the data moving between the workstation and the examination web server? Encryption schemes and security layers are not tamper proof operating over the Internet. Once the data "arrives" at the personal computer or workstation, how can the testing authority prevent the examinee from cheating through on-line collaboration, copying, looking up materials on the web, etc.? Again, test proctors usually serve this role. Finally, how can the testing authority verify the privacy and accuracy of the data the examinee provides (test responses, as well as personal information)?

Providing a secure testing environment minimizes many of these problems. That has led some organizations to build dedicated, secure test centers. Although many of the existing test-delivery vendors use secure transmission lines, a dedicated test center can also exchange data over the Internet. This *test-center-as-client* model has some obvious differences and the examinee-as-client model shown in Figure 1. First, we need to move the examinee's PC into a secure environment—usually a networked computer laboratory[3]. We additionally need to secure the software (the browser and any software that might unfairly aid the examinee) and, possibly, some of the hardware on the PC (e.g., disabling floppy or zip drives to prevent the examinee

[3] Wireless networks pose a interesting challenge to secure the network environment.

from copying test materials). Letting the examinees bring their own laptops into a testing laboratory for testing purposes is risky, regardless of the precautions taken. Second, we need to secure the Internet connections between the local network (usually handled by a file server) and the PC workstations and between the local network and the examination web server. There are technical ways of "tunneling" through the Internet using a combination of high-security encryption methods and "black boxes" to communicate between the examination file server and the PC workstations connected to a local area network (LAN) at the test center. Third, we need to put proctors in the test center (human and/or electronic surveillance) to monitor the examinees. In high-stakes testing, there seems to be no good replacement for using proctors or surveillance equipment.

Figure 2 depicts a possible configuration for LAN-based web-testing at a test center. The local area network (LAN) is comprised of a number of workstations (denoted by the circled letter, A) connected by an Ethernet (or any other LAN configuration) to a LAN file server (B). The "black box" (C) is essentially a combined *firewall*[4] and high-tech encoding equipment that encrypts and decrypts the information going out from the LAN and coming back in. The black box sends and receives the encrypted data to the web server (D). An external Internet provider may provide the server or the test center may have its own web server. Only encrypted data travels back and forth between the web

[4] A firewall is typically used to protect a web server or client from external hackers.

server on the client side and the web server on the examination server side. At the other end of the connection, another web server (E) receives and sends encrypted data to and from another black box (F). That black box "handshakes" with the first black box (C) encrypts outgoing information and decrypts incoming information. The second black box also exchanges decrypted information with the examination file server and the database system (G), neither of which is directly connected directly to the web. There are many other possible design variations. This basic model merely illustrates how a dedicated test center network can use the Internet for transmitting and receiving test-related information with a centralized examination server.

Still, dedicated test centers have serious drawbacks, too. Limited seating capacities at the centers, sometimes inconvenient locations, and scheduling complications when examinees from different test programs must compete for prime times and locations, all conspire to reduce much of the inherent flexibility and efficiency gains of WBT. Furthermore, dedicated test centers change for "seat time", which can substantially increase test delivery costs.

In short, a test-centered Internet delivery model is a hybrid of WBT that moves the entire enterprise into a secure and highly controlled environment. The Internet merely becomes a convenient transmission medium for moving information between the test center and the central examination server. There are obvious costs, but the gains in security are often worth it to testing agencies.

Over time, I would hope that alternative testing enterprise models (e.g., "plug-and-play" test centers set up at university or school-based computer laboratories) will emerge that provide all of the necessary security and other relevant test standardization features without the high overhead and facilities costs of dedicated test centers.

Measuring Complex Skills And Problem-Solving Tasks On The Web

In practice, on-line assessments vary enormously in scope and quality. Some web-based testing programs like the College Board's ACCUPLACE® are large-scale enterprises that use relatively sophisticated test delivery mechanisms like computer-adaptive testing. However, many on-line tests are low-stakes applications intended to provide practice tests or otherwise supplement on-line courses and other distance education initiatives.

There are two classes of software products available for web-based testing. One class of products includes *low-stakes* web-test authoring-and-compiler programs designed to complement the large volume of on-line courseware and distance education training projects underway around the world. This class of products includes a plethora of WBT authoring tools and HTML/script compilers. Some of these WBT tools are public domain utilities like Flashlight© and Hot Potatoes©; others are more elaborate software packages that require both substantial start-up costs and less-than-trivial per-user licensing fees. Examples include names such as LXR*TEST™, WebCT™, BlackBoard™, Question*mark* Perception™, Top Class™. Some of these require dedicated Internet servers and special security layers; others can be published to any web server on multiple platforms. Some support the new IMS Question and Test Interoperability (QTI-XML) specifications (IMS Global Learning Consortium, 2000); others use proprietary data structures, obfuscation, and encryption layers for transmittal across the web.

Another class of products includes **dedicated, custom-built test-drivers** that are more applicable to high-stakes examinations. Many of the large testing organizations and test delivery vendors either have or are building custom testing systems that fall into this class. These test driver products typically have the capability to move data using Extensible Markup Language (XML) structures, have any variety of encryption and obfuscation layers to protect both the test and response data, and employ custom browsers and rendering engines that employ client-side "plug-ins" and server-side component software to add functionality to the test. Because many of these latter products only operate within a particular company's dedicated test-center environment, their general functionality on web-based client machines and networks outside their own network is unknown.

Despite the differences in applications, one commonality of most current test production systems is that the question types and response formats tend to be similar across many of the available web-test generation products. Most web-based test authoring tools include traditional item response types such as multiple-choice items, multiple-response and extended-matching item, fill-in-the-blank and constructed response items, extended text essays, and items with hot spots (i.e., clicking on polygon area, superimposed over an image). Some of the new production tools are adding simple drag-and-drop capabilities, as well. Stimulus materials usually include text displays (with or without hyperlinks), graphics, sound clips, and multimedia with video, depending on the storage policies of the examination server sponsor and additional restrictions the Internet service provider may impose on the examination provider.

Very few, if any, test vendors support complex simulations and immersion-based, problem-solving tasks. An example of a simulation would be work-sample exercise such as solving a complex tax reporting case for a corporation using spreadsheets, research corporate financial records and conducting on-line research of the U.S. federal tax regulations. Another example would be managing a virtual medical patient. Some of these types of tests are offered for privately owned, LAN-based testing networks that have dedicated data distribution channels (e.g., Thompson Prometric and NCS Pearson). However, web-based testing has not followed this trend.

There are many excuses that could be offered as to why simulations and problem-based performance exercises are not well supported by current web-based testing applications. One of logical reason is that the Internet presently does not have adequate bandwidth for complex, data-intensive interactions between the examinee and a web server. There is a good deal of truth to that excuse. Realize most of the "web transactions" between examinees and examination web servers involve simple browser dialogs, where the server pushes web pages and scripts to the examinee's PC and the examinee "submits" response data back to the web server. The Internet time lags and the sluggish response rate of most browsers simply cannot support intensive real-time transactions. Another reason is that the customized software components usually needed to run simulations and complex performance exercises are difficult to "plug-in" to some browsers, can increase load times at the examinees' workstations, and may even compromise the security of the examination system if the component software is actually downloaded to a workstation. Hopefully, those current limitations will not hamper future developments. A final reason may simply be that the companies creating the web-based software do not understand the full scope of assessment needs. Some education seems in order.

In any case, technological limitations are not good for web-based testing. If we force our assessments to adapt to the lowest common denominator—in this case rather restrictive tests based solely upon multiple-choice and limited response technology—web-based testing will have added little except to make quizzes, surveys and practice examinations that use those formats a bit more convenient for examinees to access. The simple fact is that current web-based test drivers tend to be limited by their [necessary] reliance on current browser technologies.

The next generation of web-based test drivers will need to make serious design improvements that flexibly manage complex item types, problem-solving tasks, and performance-based exercises. It is not currently clear whether the IMS Question and Test Interoperability specifications (IMSWP-1 Version A) for extensible markup structures—i.e., QTI-XML (IMS Global Learning Consortium, 2000) will address these needs.

Perhaps some of the more promising areas of potential development for web-based testing will evolve by making more extensive use of Microsoft's new release of Visual Basic.NET, Windows.NET®, and the Server Explorer; as well as new developments with the Object Model for ActiveX Data Objects™ and Active Server Pages. Some would call many of these enhancements new versions of Applications Program Interfaces (APIs). In any case, their potential for integrated development of distributed, web-based applications seems promising. Other work seems needed to develop test drivers that interact more strongly with "middleware" APIs and server-side components running on multi-tiered server platforms to speed up behind-the-scenes processing, allow more intensive transactions with examinees and client servers, and to better secure the data. Transaction-intensive like adaptive testing and simulations will especially demand these capabilities.

We also need to keep in mind that the **production demands** for new test items, including simulations and problem-solving performance exercises, can be enormous under computer-based testing (CBT). Web-based testing is no different. Having large item pools can help to mitigate security risks due to examinees memorizing and sharing items. Unfortunately, creating large item pools is a nontrivial enterprise and implies that testing agencies must engage in large-scale item authoring efforts to mass produce as many high quality items as possible. Authoring tools are needed to support these efforts. Most of the major authoring tools include "templates" for standard multiple-choice and constructed response item types. The templates provide "blanks" that hold the content of the items (e.g., stem, exhibits and distractors for a multiple-choice item). The item author fills in the blanks and the authoring software renders the HTML code and associated scripts that will be executed by the test driver (browser) when the item is used. However, moving toward using complex, computerized performance assessments means developing more complicated templates and training item writers to effectively use those tools.

<u>Integrating Modern Item Selection And Test Assembly Algorithms</u>

With some exceptions, the majority of web-based testing programs are extremely weak when it comes to their actual test assembly capabilities. There seem to be four different perspectives on what constitutes "item selection" for a computer-based test or web-based test.

One perspective is that the test designer should manually build each test form. In our computerized world, this implies that the test designer needs tools to query and drag test items, one-by-one, onto a "test form". The test assembly component of the software therefore builds a list of items and publishes the items in HTML or some other format. If the software is highly sophisticated, it may allow the test designer to specify that items in the list should be randomly scrambled and, possibly, that the multiple-choice distractors should be randomly scrambled by running a script when the item is rendered.

A second perspective is that item selection is simply the process of randomly selecting items from an item pool. Unfortunately, this practice often leads to tests of differing difficulty and sometimes, different content. Reliability and validity could become seriously compromised. Even more unfortunate is the fact that random item-selection is one of the primary "benefits" often touted in the marketing materials for certain web-based testing products.

A third perspective is that item selection should be adaptive, where selecting items tailored to the proficiency level of the examinee maximizes statistical precision of the test or allows the test length to be reduced for some examinees without changing the level of score precision across examinees. Because it uses item response theory (IRT), adaptive testing has the advantage of being able to "calibrate" an examinee's performance to a common score scale, even if their particular test form was easier or harder than other examinees' tests.

Computer-adaptive testing (CAT) has been around since the 1970's and many organizations have embraced the technology for tests administered over dedicated networks. Web-based adaptive testing (WBAT) has been used less frequently. Thanos, Way and Elliot (2000) described a particular application of WBAT for an algebra test. The Medical Council of Canada's Qualifying Examination (Part I) and the Clinical Reasoning Skills Examination (MCC, 2000) are two of the first high-stakes examples of WBAT implementations. Both examinations adaptively administer "testlets" (clusters of items with a predefined content balance and case-based item sets) via a dedicated, secure web-based testing system that uses the Internet to link computer laboratories in Canadian medical schools with a central examination web server. Sitting in a proctored testing laboratory, the examinee completes and submits a testlet from his or her PC. The examination server scores and aggregates all previous performance and selects the next testlet using an adaptive algorithm. The new testlet is published and pushed to the examinee's PC as web pages.

Computer-adaptive testing (CAT) is a theoretically sound idea that has had its difficulties in practice, especially for "on-demand" testing on a continuous basis. Large item pools and sophisticated "exposure control" algorithms are usually needed to reduce the likelihood that examinees will see particular segments of the item pool. Also, some have argued that the typical item selection algorithms used in CAT ignore content and other features important to establishing test validity.

A final perspective on item selection, and perhaps the most ambitious in scope, is that all computer-based tests should embrace automated test assembly (ATA) algorithms and heuristics to ensure that both statistical balance and content balance can be consistently achieved on every test form. Hundreds or even thousands of test-content and other attributes can be introduced as "constraints" to be met. These mathematical algorithms and heuristics can also select items, intact testlets or even test forms to meet the same or varied difficulty levels (Luecht & Nungester, 1998). It is even possible to integrate ATA procedures as part of the runtime web application to replace random item selection or CAT algorithms[5]. Many of the major testing companies now employ ATA software for their computer-based and paper-and-pencil tests. Unfortunately, ATA technology seems virtually unknown in many web-based testing circles, especially those that support distance education and other low-stakes examination needs. Once again, some education seems in order.

There is one drawback to CAT and ATA for web-based testing. Unless the tests are pre-constructed (see, for example, Luecht, 1998; Luecht & Nungester, 1998), real-time CAT and ATA applications can add computational overhead at the [web] server level. It is difficult to predict how those loads will impact lag times and overall performance in a typical web-based application. ATA operating as "middleware" running on multi-tier server platforms may help alleviate some of the overhead and processing time, especially for high volume applications.

In any case, web-based testing developers need to realize that there is a sophisticated science behind modern testing—a science that includes more than just randomly selecting items. The Association of Test Publishers (ATP) recently issued its revised *Guidelines for Computer-Based Testing* (ATP, 2001) that specifically address some of these issues and provide recommendations for developing, validating, and implementing computer-based tests. Information specific to new testing technologies such as adaptive testing, linear-on-the-fly testing, applications of "testlets", and automated test assembly are included.

Storing And Processing Relevant Examinee Response Data

Many commercial web-based testing applications consider it part of their "service" to provide the end-users with test scores and reports showing limited statistical item performance measures (mean difficulties, frequencies of various responses, and possibly, item-test correlations). Unfortunately, what most web-testing end-users ought to demand is the raw data. Raw response data is essential for conducting item analyses, calibrating items using an IRT model, or

[5] CAT algorithms are heuristics. ATA heuristics and algorithms merely expand the number of possible constraints and objective functions

for conducting various types of cheating analyses that might detect examinees collaborating on the Web, or simply for psychometric research. Furthermore, process information such as timing data and when/how often various components or on-line resources were used during the test are relevant information from a measurement perspective. At the very least, those types of measures can help in research and can further provide empirical evidence when investigating cheating and other aberrant response cases.

It is interesting to note that a review of the QTI-XML specifications (IMS Global Learning Consortium, 2000) does not appear to include any explicit recommendations with respect to storing raw response-level data, timing data or other process variables. Scoring is viewed as something handled by embedded scripts, yielding a numeric quantity for each item (i.e., correct = 1; incorrect = 0).

The good news is that the Extensible Markup Language (XML) on which QTI-XML is based is general enough to incorporate almost any logical data structures and data types. There are potential overhead costs, in terms of storage space and system performance degradation, if too much information (encapsulated in XML-structured results files) is sent from a testing workstation to the examination web server. Nonetheless, it becomes a political or financial rather than a technical decision as to which information is retained and under what conditions. For example, one might want to retain timing data and other process information during experimental pretesting of new items and turn off that feature for operational test items. The web-based testing application ought to be able to generate the raw data, regardless of whether or not it is used all the time. The inability of most of the current commercial WBT products to routinely provide this type of data is a grave limitation and renders many of those products virtually useless for most serious testing applications and psychometric research. We can only hope that those commercial organizations producing WBT software will modify their software.

Large-Scale Distribution Of "High-Bandwidth" Tests

The bandwidth dilemma is obviously serious for web-based tests that include high-density photographs, video clips, audio files, or any data intensive components. Data streaming/paging and compression technologies have improved enormously, however, there is still serious degradation of performance when large data files or large amounts of digital data need to be transported over the Web. Perhaps the only reasonable solution will be to wait for WWW2 (the World Wide Web 2) and its greatly enhanced bandwidth.

Before moving on to the final challenge, it is perhaps worth mentioning one interesting development in the area of data streaming that comes from a joint venture between Warner Brothers Studios, a Hollywood Film Company, and TRW, a high-tech, California-based defense industry firm. The PicturePipeline™ can move full motion video—in real-time and fully encrypted using DES 128-bit encryption no less—over the Internet. The current application of the

PicturePipeline is limited to being able to distribute digital movies and such for multi-channel editing. However, many other applications, including web-based testing, could benefit from that type of capability to securely and quickly transport high-bandwidth data over the Web.

Optimal Ergonomic Design Of Web-Based Testing Interfaces

Despite the guidelines and "best practices" principles alluded in my introduction, many WBT software products are developed and marketed with limited or no "usability" research conducted, much less impact studies. Web-based testing software vendors almost routinely post "IMS Compliant" on their websites but fail to list "ADA Compliance" (Americans with Disabilities Act) or other information suggesting that their products follow best practices and adhere to the *1999 Standards for Educational and Psychological Testing* or the ATP *Guidelines*. The simple fact is that there is no certifying body for sound software development, especially in critical applications like testing. Maybe it is time for a change. In any event, it is time for more research.

The field of human factors has provided important research and guidelines concerning the proper ergonomic design of many products in other areas. During the 1980's the Human Factors Society conducted many studies on interface designs covering research topics ranging from color and font selections to menu design. It seems reasonable to demand more of that type of research for computer-based and web-based tests. Further, more than "consumer preference" needs to drive that research. How much training is needed to learn to effectively use keyword search engines? How can "help" systems be designed to minimize the time the examinee spends searching irrelevant information? How much training, on average, do examinees in the target population need to master the various item types used? Do certain interface designs or components facilitate performance? Does others penalize particular individuals in a manner that is irrelevant to the purpose of the test?

Much of this research should also be combined with psychometric research to directly assess its impact pacing and performance. Ultimately, regardless of the purpose and use of a web-based assessment, we need high-quality, fair, and efficient measurement instruments. Research can help to make that happen.

Conclusions

It may be that this paper has raised many questions and provided few answers. I hope that the ideas are stimulating and concrete enough to offer some suggestions about new directions for research. I further hope that I was able to justify the need for better education across many sectors of the web-based testing community about technical aspects of psychometrics and high-stakes testing needs. Finally, I would hope that I adequately conveyed the message that there needs to be adherence to standards and principles of professional practice and science.

References

Association of Test Publishers. (2001). *Guidelines for Computer-Based Testing*. (Author).

Green, K. C. (2000, October). *The 2000 National Survey of Information Technology in Higher Education*. Encino, CA: Center for Educational Studies, The Clairmont Graduate University.

IMS Global Learning Consortium, Inc. (2000). IMS Question and Test Interoperability Specification. (Author).

Luecht, R. M. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement, 22*, 224-236.

Luecht, R. M. & Nungester, R. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*.

Lyman, H. B. (1998). *Test Scores and What They Mean, 6th Edition*. Boston: Allyn and Bacon.

Medical Council of Canada. (2000). *MCC Qualifying Examination (Part I). Clinical Reasoning Skills Examination*. (Author).

Thanos, P., Way, W. D. & Elliot, S. (2000, April). *Developing a research plan for an online assessment program*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Wall, J. E. (2000). Technology-delivered assessment: Power, problems, and promise. In *Cybercounseling and Cyberlearning: Strategies and Resources for the Millenium*.

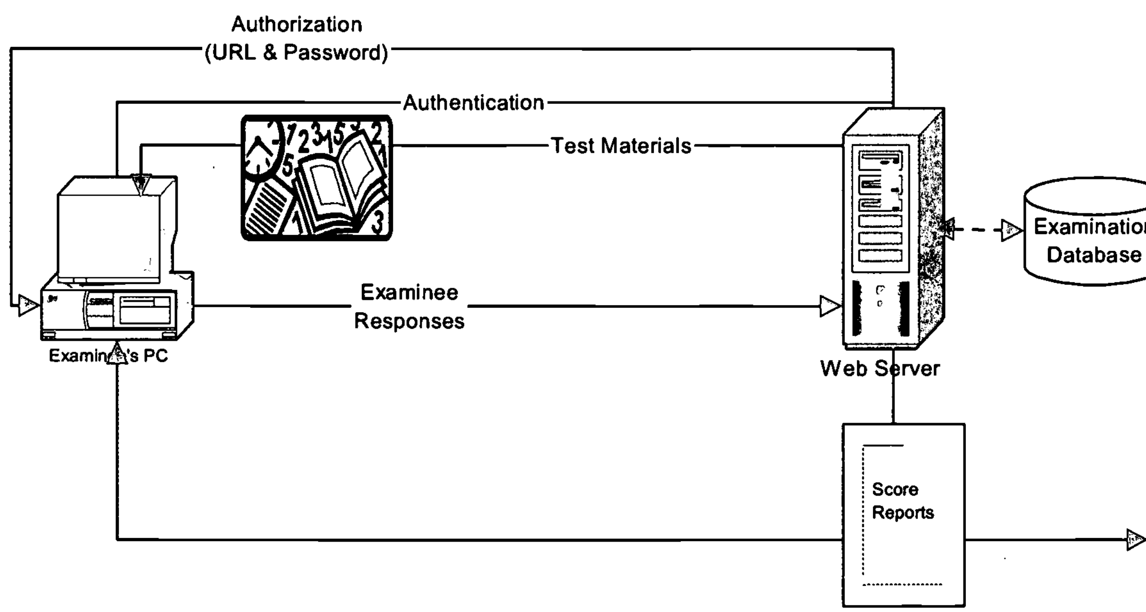

Figure 1. Conceptual Diagram of a Test Delivered via the Internet

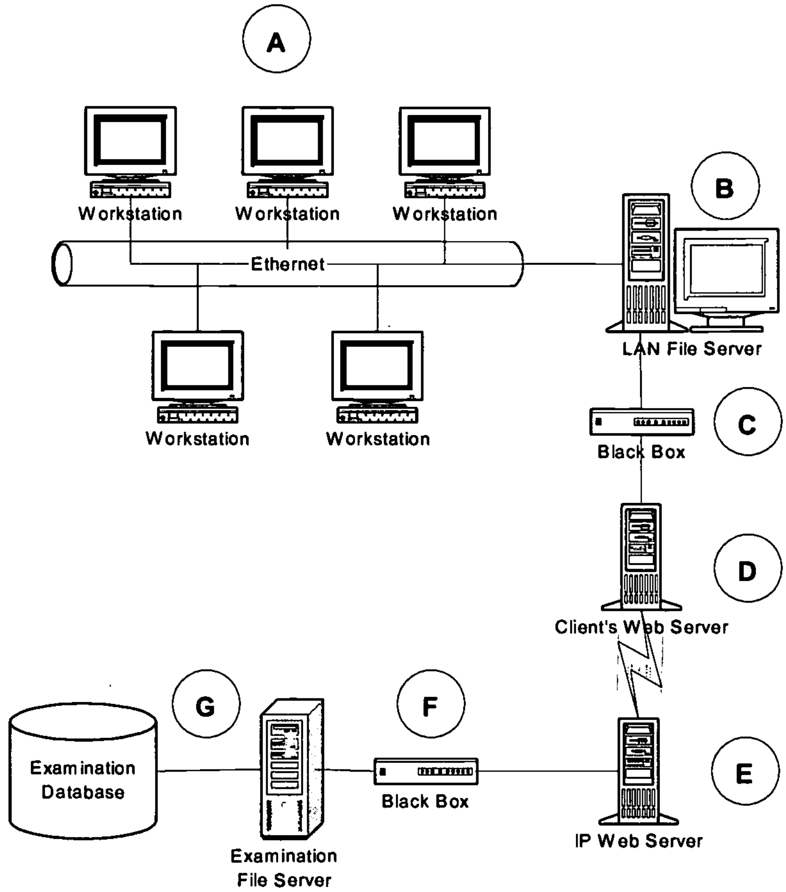

Figure 2. Diagram of a Secure "Tunnel" Connect between a LAN-based Testing Center and an Examination File Server

AERA

ERIC®

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM033467

I. DOCUMENT IDENTIFICATION:

Title: Challenges of Web-Based Assessment

Author(s): Richard M. Luecht

Corporate Source: University of North Carolina at Greensboro

Publication Date: April 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
XX		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Richard M. Luecht	Printed Name/Position/Title: Richard M. Luecht	
Organization/Address: 209 Curry, Box 26171 Greensboro, NC 27402-6171	Telephone: 336-334-3473	FAX:
	E-Mail Address: rmluecht@uncg.edu	Date: 11/16/01

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com